

.opendata to DIF mapping

Derived from <http://project-open-data.github.io/metadata-resources/#common-core-required-fields-equivalents>

See general Schema documentation at <http://project-open-data.github.io/schema/> and mapping to ECHO10 schema at [.opendata to ECHO10 mapping](#)

NOTE: in XSLT 1 (at least using xsltproc on the command line), the xpaths must have the dif namespace prefix (`xmlns:dif="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/"`). However, in Ruby, Nokogiri lets us include a prefix wildcard in xpaths ('`*`') which will match any namespace prefix. We use this syntax where possible.

“Common Core” Required Fields

Field	Definition	DIF Mapping
title	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	<code>/*:Entry_Title</code>
description	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	<code>/*:Summary/*:Abstract</code>
keyword	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.	<code>/*:Keyword</code> (multiple)
modified	Most recent date on which the dataset was changed, updated or modified.	<code>/*:Last_DIF_Revision_Date</code>
publisher	The publishing entity.	<code>/*:Data_Set_Citation/*:Dataset_Publisher</code>
contactPoint	Contact person's name for the asset.	<code>/*:Personnel/*:First_Name /*:Personnel/*:Last_Name</code> (<code>/*:Personnel/*:Role</code>)
mbox	Contact person's email address.	<code>/*:Personnel/*:Email</code> or support@earthdata.nasa.gov if this returns no value.
identifier	A unique identifier for the dataset or API as maintained within an Agency catalog or database.	<code>echo_collection_id</code>
accessLevel	The degree to which this dataset could be made publicly-available, <i>regardless of whether it has been made available</i> . Choices: public (Data asset is or could be made publicly available to all without restrictions), restricted public (Data asset is available under certain use restrictions), or non-public (Data asset is not available to members of the public)	'public'

“Common Core” Required-if-Applicable Fields

Field	Definition	DIF Mapping
bureauCode	Federal agencies, combined agency and bureau code from OMB Circular A-11, Appendix C in the format of 015:11.	'026:00' <i>NOTE: in some cases a different bureau code(e.g. USGS) may be more appropriate. When these are identified, we will modify the mapping.</i>
programCode	Federal agencies, list the primary program related to this data asset, from the Federal Program Inventory . Use the format of 015:001	'026:001' <i>NOTE: in some cases a different program code may be more appropriate. When these are identified, we will modify the mapping.</i>

accessLevelComment	An explanation for the selected “accessLevel” including instructions for how to access a restricted file, if applicable, or explanation for why a “non-public” or “restricted public” data asset is not “public,” if applicable. Text, 255 characters.	nil
accessURL	URL providing direct access to the downloadable distribution of a dataset.	nil
webService	Endpoint of web service to access dataset.	nil
format	The file format or API type of the distribution.	nil
license	The license with which the dataset or API is published. See Open Licenses for more information.	nil
spatial	The range of spatial applicability of a dataset. Could include a spatial region like a bounding box or a named place.	//*:Spatial_Coverage/*:Westernmost_Longitude, //*:Spatial_Coverage/*:Southernmost_Latitude, //*:Spatial_Coverage/*:Easternmost_Longitude, //*:Spatial_Coverage/*:Northernmost_Latitude
temporal	The range of temporal applicability of a dataset (i.e., a start and end date of applicability for the data).	//*:Temporal_Coverage/*:Start_Date'//*:Temporal_Coverage/*:Stop_Date

Expanded Fields

Field	Definition	DIF Mapping
theme	Main thematic category of the dataset.	nil
dataDictionary	URL to the data dictionary for the dataset or API. Note that documentation other than a data dictionary can be referenced using Related Documents as shown in the expanded fields.	nil
dataQuality	Whether the dataset meets the agency's Information Quality Guidelines (true/false).	nil
distribution	Holds multiple download URLs for datasets composed of multiple files and/or file types	URL to search this collection in reverb
accrualPeriodicity	Frequency with which dataset is published.	nil
landingPage	Alternative landing page used to redirect user to a contextual, Agency-hosted “homepage” for the Dataset or API when selecting this resource from the Data.gov user interface.	<p>URL to search this collection in reverb.</p> <p><i>NOTE: The detailed API description says</i></p> <div style="border-left: 1px solid black; padding-left: 10px;"> "This allows agencies to better specify what a visitor receives after selecting one of the agency's datasets on Data.gov or in third-party mashups." </div>
language	The language of the dataset.	'en-US'
PrimaryITInvestmentUII	For linking a dataset with an IT Unique Investment Identifier (UII)	nil
references	Related documents such as technical information about a dataset, developer documentation, etc.	//*:Related_URL/*:URL (multiple)
issued	Date of formal issuance.	nil

systemOfRecords	If the systems is designated as a system of records under the Privacy Act of 1974, provide the URL to the System of Records Notice related to this dataset.	nil
-----------------	---	-----

Simple conversion stylesheet

Note that the catalog-rest code does not use this stylesheet. This is only meant as a testing tool to see what values catalog-rest should be pulling based on the above mapping.

Run the following on the command line after saving the stylesheet to a file to apply it to a metadata file:

```
xsltproc <XSLT File> <Metadata File> | xmllint --format -

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0" xmlns="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/" xmlns:dif="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/ http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/dif_v9.8.4.xsd">

<xsl:template match='/'>
<opendata>
<title>
<xsl:value-of select="//dif:Entry_Title"/>
</title>
<description>
<xsl:value-of select="//dif:Summary/dif:Abstract"/>
</description>
<keywords>
<xsl:for-each select="//dif:Keyword">
<keyword><xsl:value-of select=". /></keyword>
</xsl:for-each>
</keywords>
<modified>
<xsl:value-of select="//dif:Last_DIF_Revision_Date"/>
</modified>
<publisher>
<xsl:value-of select="//dif:Data_Set_Citation/dif:Dataset_Publisher"/>
</publisher>
<contactPoint>
<xsl:value-of select="//dif:Personnel/dif:First_Name"/>
<xsl:text> </xsl:text>
<xsl:value-of select="//dif:Personnel/dif:Last_Name"/> <xsl:text> </xsl:text> (<xsl:value-of select="//dif:Personnel/dif:Role"/>)
</contactPoint>
<mbox>
<xsl:value-of select="//dif:Personnel/dif:Email"/>
</mbox>
<spatial>
<xsl:value-of select="//dif:Spatial_Coverage/dif:Westernmost_Longitude"/>, <xsl:value-of select="//dif:Spatial_Coverage/dif:Southernmost_Latitude"/>, <xsl:value-of select="//dif:Spatial_Coverage/dif:Easternmost_Longitude"/>, <xsl:value-of select="//dif:Spatial_Coverage/dif:Northernmost_Latitude"/>
</spatial>
<temporal>
<xsl:value-of select="//dif:Temporal_Coverage/dif:Start_Date"/>/<xsl:value-of select="//dif:Temporal_Coverage/dif:Stop_Date"/>
</temporal>
<references>
<xsl:for-each select="//dif:Related_URL">
<reference><xsl:value-of select='dif:URL' /></reference>
</xsl:for-each>
</references>
<opendata>
</xsl:template>
</xsl:stylesheet>
```